THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 261, No. 23, Issue of August 15, pp. 10797-10801, 1986
Printed in U.S.A.

Nucleotide Sequence of the Gene for the Ferrienterochelin Receptor FepA in *Escherichia coli*

HOMOLOGY AMONG OUTER MEMBRANE RECEPTORS THAT INTERACT WITH TonB*

(Received for publication, March 13, 1986)

Michael D. Lundrigan‡ and Robert J. Kadner§

From the Department of Microbiology, University of Virginia School of Medicine, Charlottesville, Virginia 22908

We have determined the nucleotide sequence of the *Escherichia coli fepA* gene, which codes for the outer membrane receptor for ferrienterochelin and colicins B and D. The predicted FepA polypeptide has a molecular weight of 79,908 and consists of 723 amino acids. A 22-amino acid leader or signal peptide preceded the mature protein. With respect to overall composition, hydropathy, net charge and distribution of nonpolar segments, the FepA polypeptide was typical of other *E. coli* outer membrane proteins, except that FepA contained 2 cysteine residues. Comparison of the deduced amino acid sequence of FepA with that of three other TonB-dependent receptors (BtuB, FhuA, and IutA) revealed only a few regions of sequence homology; one of these included the amino-termini. An amino acid substitution within the conserved amino-terminal region of BtuB resulted in production of a receptor that had normal binding functions but was incapable of energy-dependent transport of vitamin B_{12}. This result suggests that the amino-terminal end of these four polypeptides is involved in interaction with the TonB protein or another step of energy transduction. Three other regions of homology were shared among the four proteins: one near residues 50 to 70, another at about residue 100 to 140, and the last between 20 and 40 amino acid residues from the carboxyl terminus. The function of these three regions remains speculative.

At physiological pH iron is sequestered as insoluble complexes and therefore is not readily available to cells. Many organisms, including *Escherichia coli*, extract iron from the environment by synthesizing and excreting iron chelators known as siderophores (for reviews see Refs. 1 and 2). *Escherichia coli* normally produces the siderophore known as enterochelin or enterobactin. Enterochelin is a cyclic trimer of 2,3-dihydroxy-*N*-benzoyl-L-serine and its synthesis is regulated by the intracellular iron supply.

In addition to making enterochelin, iron-starved cells produce a number of proteins some of which are directly involved in the transport of iron chelates (2). FepA is an 81,000-dalton *E. coli* outer membrane protein that functions in the initial step of iron uptake by binding ferrienterochelin (3). The subsequent transport of ferrienterochelin across the outer membrane, presumably into the periplasmic space, requires TonB function, which seems to provide energy for a number of outer membrane receptor-dependent processes. The specific steps of iron transport from the periplasm into the cytoplasm are unclear but involve the products of the *fepB* and *fepC* genes[1] (4), probably in uptake across the cytoplasmic membrane, and the protein product of *fes* (ferrienterochelin esterase) which cleaves ferrienterochelin to allow release of its iron (5). Another outer membrane protein produced by wild-type *E. coli* under iron stress is FhuA, a 76,000-dalton protein that serves as the receptor for the hydroxamate siderophore ferrichrome, colicin M, and phages T5 and ϕ80 (6). Although desferri-ferrichrome is produced by the fungus *Ustilago sphaerogena*, it can be used by *E. coli* as an iron carrier. Strains of *E. coli* harboring the ColV plasmid excrete the siderophore aerobactin and produce an outer membrane receptor, IutA, responsible for its binding and transport (7, 8). As in the enterochelin system, both FhuA and IutA require TonB function for the energy-dependent step of shuttling iron across the outer membrane; both of these systems require different genes (*fhuBCD*) for the subsequent steps of uptake (9).

Transport of vitamin B_{12} (cyanocobalamin) is analogous to the uptake of iron in that a TonB-dependent outer membrane receptor and additional inner membrane proteins are required (2, 10). The vitamin B_{12} outer membrane receptor, called BtuB, also serves as the receptor for the lethal agents, bacteriophage BF23 and the E colicins (11). Whereas transport of vitamin B_{12} requires TonB function, the binding to BtuB by the vitamin, as well as entry of the lethal agents, does not. The suggestion by Heller and Kadner (12) that the amino-terminal region of BtuB interacts with TonB provided the impetus for this study since the domains of these receptors that interact with TonB might have homologous amino acid sequences.

EXPERIMENTAL PROCEDURES

Plasmids, Bacterial and Phage Strains—Bacteriophages M13mp9, mp8, mp19, mp18 (13), tg130, and tg131 (14) and their host JM101 were from laboratory stock. Plasmid pPC104, described in Fig. 1, is a derivative of pBR322 carrying an 8-kilobase *SalI-EcoRI* fragment able to complement *entD*, *fepA*, and *fes* mutations (15). It and the *fepA* deletion strain, UT6900, were kindly provided by Charles Earhart (University of Texas at Austin).

Genetic Techniques—Plasmid DNA for subcloning *fepA* was obtained from chloramphenicol-amplified cultures by the method of Katz *et al.* (16) and purified by CsCl-ethidium bromide equilibrium centrifugation. M13 replicative form and template DNA was obtained as recommended by the manufacturer of the cloning and sequencing kits (Amersham Corp.). Enzymatic digestions and ligations of DNA were according to the manufacturer's (Bethesda Research Laborato-

* This work was supported by Grant GM19078 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of National Institutes of Health Training Grant CA09109.

§ To whom correspondence should be addressed.

[1] C. F. Earhart, personal communication.

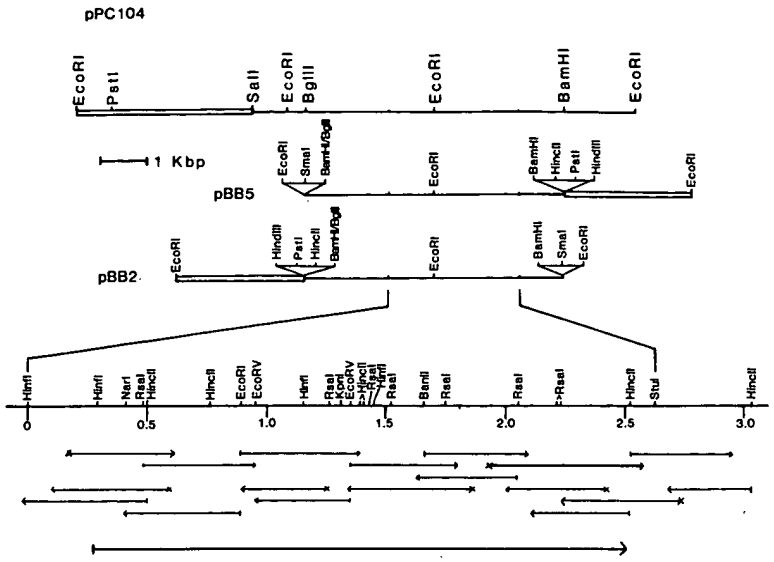

FIG. 1. **Sequencing strategy and restriction map of the *fepA* gene.** The plasmids from which DNA fragments were obtained are shown at top. *Open bars* represent vector DNA. The direction and extent of a sequence reading is shown by the *small arrows* under the restriction map. *Arrows* starting with an *X* indicate fragments obtained by exonuclease III digestion, whereas *arrows* which begin with a *bar* represent fragments obtained by restriction endonuclease cleavage. The *large arrow* at the bottom designates the FepA coding region and the direction, although not the extent, of transcription.

ries or New England Biolabs) recommendations. Transformation or transfection of bacterial cells with DNA was done by heat shocking a mixture of DNA and cells that had been made competent by suspension in 0.1 M calcium chloride (17).

Nucleotide sequence determination was performed by the Sanger dideoxy chain-terminating technique (18) from M13 templates using deoxyadenosine 5′-([α-^{35}S]thio)triphosphate as described in the Amersham sequencing kit. The computer program described by Staden (19) was used to analyze sequence data.

Exonuclease III digestion was used to generate *fepA* deletions by a method similar to that reported by Henikoff (20). The procedure involved the initial subcloning of *fepA* as a 5.4-kilobase *Bgl*II-*Bam*HI fragment into pUC8. Subclones with *fepA* in opposite orientations were digested with *Pst*I and *Sal*I and then with exonuclease III. Since *Eco*III initiates degradation of unpaired four-base 3′ ends inefficiently, digestion occurs mainly on the 5′ overhangs (in this case toward *fepA*). Blunt ends were made by treatment with S1 nuclease. The DNA was fractionated by electrophoresis on a 0.8% agarose gel and various molecular weight fractions were isolated. The ethanol-precipitated DNA was ligated with T4 ligase and transformed into UT6900.

RESULTS AND DISCUSSION

Coderre and Earhart (15) have shown that plasmid pPC104 carries the promotor and coding region for FepA. The *Bgl*II-*Bam*HI fragment of pPC104 (Fig. 1) was subcloned into the *Bam*HI site of pUC8 in both orientations. Deletion derivatives of these two plasmids were generated via exonuclease III digestion. When the deletion plasmids were ordered according to size and complementation activity, the position of the *fepA* gene boundaries could be discerned. The restriction maps of the parent plasmids and the strategy for nucleotide sequence determination are shown in Fig. 1. The position and extent of the *fepA* coding sequence is represented by the *large arrow* at the bottom of the figure. Initially, sequencing was performed on the *Hin*dIII to *Eco*RI DNA fragments from the deletion plasmids subcloned into phage M13mp8 using the dideoxy chain termination method. These data served to locate restriction sites which could provide overlapping DNA fragments carrying the opposite strand. The particular M13 vector used was dependent on the restriction fragment to be sequenced. All of the *fepA* coding sequence was determined from both strands and readings extending across all restriction sites were obtained.

The fepA Gene—Fig. 2 shows the nucleotide sequence of *fepA*, beginning at the *Hin*fI site 892 base pairs to the left (Fig. 1) of the insert's unique *Eco*RI site and extending to the *Stu*I site 1,728 base pairs to the right of the *Eco*RI site. The deduced amino acid sequence of the only open reading frame long enough to code for a polypeptide of about 81,000 molecular weight is shown above the nucleotide sequence. The direction of transcription and hence of translation is counterclockwise (*i.e.* from *lip* toward *purE*) in agreement with the results of Fleming *et al.* (21). Possible translation initiation codons are found at positions 204, 226, and 271. The third ATG is the most likely translation start site for two reasons: (i) the succeeding 21 amino acids were typical of a signal sequence present on other outer membrane proteins with the probable cleavage site after the alanine-glutamine-alanine (22); and (ii) the sequence preceding this ATG had greater complementarity to the 3′ end of 16 S ribosomal RNA (Shine-Dalgarno sequence) than did the corresponding regions of the other two.

Upstream from the translation initiator site are a number of potential promotor -35 and -10 regions. The sequence TGACTGCGT, starting at position 160, also occurs upstream of *fhuA* and *btuB* (see Miniprint Section).[2] Although a typical -10 region does not follow 16–19 base pairs downstream, this sequence may be the -35 region for these genes or it may be a recognition site for a common regulatory effector molecule. *btuB* expression has been observed to be regulated by iron,[3] as are *fepA* and *fhuA* (3, 23). Therefore this sequence could be the binding site for an iron repressor protein such as *fur* (23). A less conserved sequence occurs upstream of *iutA* but it is uncertain whether this gene would have the same regulatory controls since *iutA* is part of the aerobactin operon and its promotor is far upstream of the *iutA* gene (24).

The UGA termination codon is followed by a region in which either of two stable stem and loop RNA structures can form. These inverted repeats are indicated by *arrows* in Fig.

[2] Portions of this paper (including Tables 1–3 and an additional Fig. 1) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 86M-0806, cite the authors, and include a check or money order for $1.20 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

[3] K. Heller, personal communication.

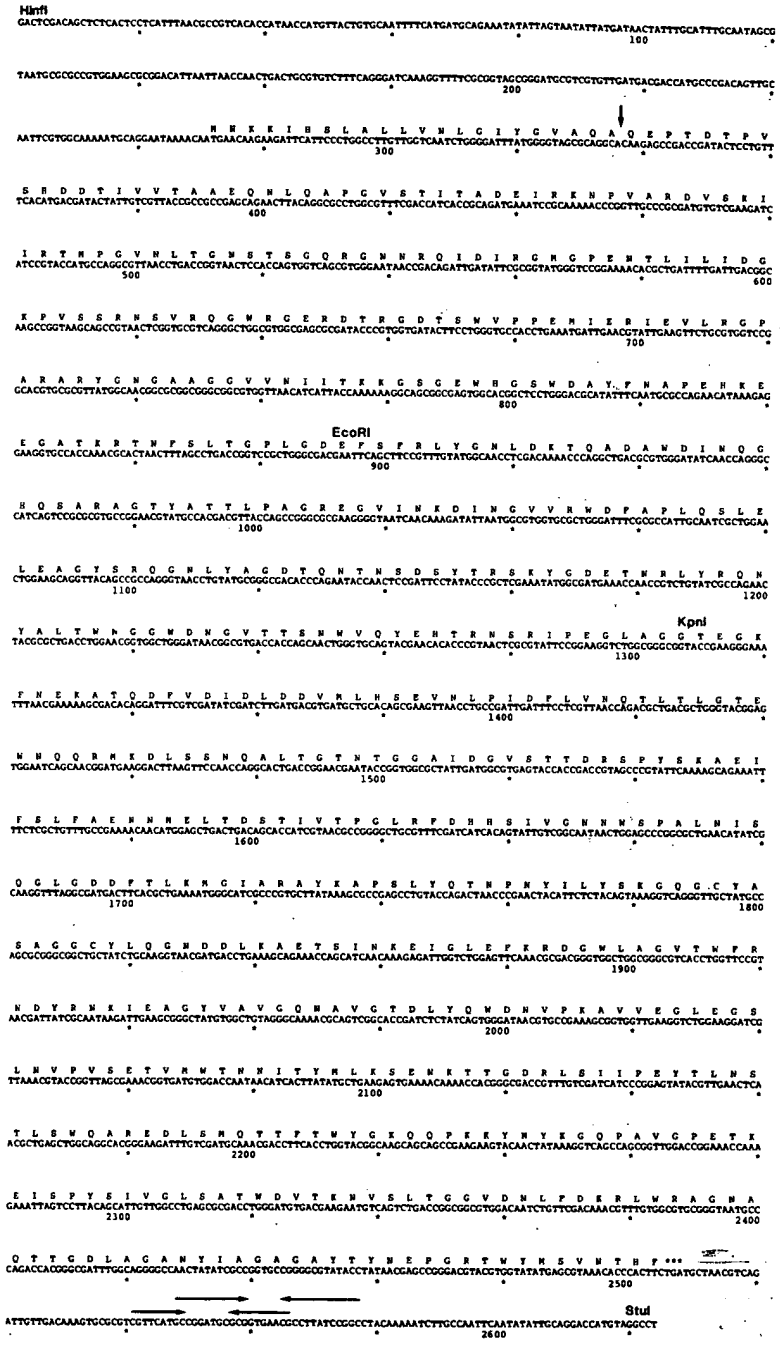

FIG. 2. **Nucleotide sequence of the *fepA* gene and the deduced amino acid sequence of the FepA precursor.** The antisense strand is shown with the amino acid sequence, in one-letter code, above it. Representative restriction sites are indicated by the names above their respective recognition sequence. The *vertical arrow* is the putative signal sequence cleavage site. *Horizontal arrows* designate a region of dyad symmetry following the termination codon.

2 and the structures they can form are presented in the Miniprint Section. Because neither of these structures are followed by a run of Ts, as is the case for rho-independent terminators, it is likely that this region is a rho factor-dependent transcription terminator for the *fepA* gene. The rho-dependent terminators for λt_{R1} and tRNA-Tyr are also deficient in Ts at the 3′ ends, whereas terminators in which a stem loop structure is followed by 3 to 7 T residues can function independently of rho (25). In addition, following the stem loop structures are the sequences CAAAAA and CAATTCAA which are similar to the sequences CAAAAG and CAAUCAA found at or near the 3′ end of several rho-dependent transcripts (26).

Features of the FepA polypeptide—The predicted FepA protein consists of 723 amino acids preceded by a 22-residue leader or signal peptide. The molecular weight of the mature protein was calculated to be 79,908 in agreement with the value determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Similar to other outer membrane proteins the number of charged residues in FepA were 21% of the total. The mean hydropathy calculated according to Kyte and Doolittle (27) was −0.64; by comparison that of BtuB was −0.55. Hydrophobic amino acids were fairly evenly dispersed throughout the polypeptide as were charged residues; no uncharged regions of sufficient length to span the outer membrane as an α-helix were present. With respect to net charge, FepA (net charge of −14) was like the porins (OmpF, −11; OmpC, −14; and PhoE, −9) (28) and the TonB-dependent receptors (BtuB, −14; FhuA, −16; and IutA, −13). FepA has 2 cysteines, FhuA has 4, but the porins and the other two

TonB-dependent receptors have none. Like BtuB the distribution of tyrosine residues in FepA was not even. The amino-terminal end of FepA was devoid of tyrosines to residue 133, whereas clusters of tyrosines occurred in the carboxyl-terminal half of the protein.

Codon usage (see Miniprint Section) was typical of that of a strongly expressed protein since codons recognized by the most abundant tRNA species were preferred (29, 30). This was in agreement with the observation that the derepressed level of FepA in the outer membrane can rival the porin content (3).

Homology among TonB-dependent Receptors—Fig. 3 presents the sequences in the four regions of significant amino acid homology shared by the TonB-dependent receptors, FepA, BtuB, FhuA, and IutA. Region I includes the amino-termini of the four proteins. In cells carrying the *btuB451* mutation, vitamin B_{12} uptake activity is abolished while other receptor functions are normal, possibly owing to failure of the altered receptor to interact with TonB. The mutation in the *btuB451* allele changed the eighth amino acid of the mature BtuB from leucine to proline (12). Fig. 3 shows that this leucine residue is near the center of homology region I. The finding that the amino-termini of these four TonB-dependent receptors are homologous supports the proposal that this region is important for proper interaction with TonB or energy transduction.

Homology region II is near the carboxyl terminus and is the most highly conserved of the four homology regions but is also the shortest.

Region III is the least conserved of these sequences; however, like region IV, the homology extends for a greater length than does that of regions I or II. In positions where amino acid residues are not identical, conservative replacements were frequently observed.

Comparison of the nucleotide sequences encoding the homologous regions (not shown) revealed that the sequences were disparate, suggesting that these proteins did not arise from a common ancestor but rather resulted from convergent evolution. Functions of the homologous regions could include interaction with TonB, export and localization of the proteins to the outer membrane, interaction with lipids or lipopolysaccaride, or involvement in the ligand transport mechanism, possibly by channel formation. No substantial regions of homology among the four TonB-dependent receptors and the outer membrane proteins OmpC, PhoE, and LamB were found by the homology search program, FASTP, of Lipman and Pearson (32). Very short regions of homology between FepA and OmpF, and FepA and OmpA were found. Homology with OmpF was found slightly downstream from region IV. The region of FepA homologous to a portion of OmpA is about 70 amino acids upstream of region II. In OmpA this area of homology is slightly upstream of the region identified by Nikaido and Wu (33). Other homology regions described by Nikaido and Wu for the porins, LamB, and OmpA were not apparent in FepA. Thus these receptors can be considered as a distinct class of outer membrane proteins in which amino acid sequence homology occurs only in a few segments. Investigations of the function of these conserved regions should prove interesting.

REFERENCES

1. Neilands, J. B. (1981) *Annu. Rev. Biochem.* **50**, 715–731
2. Neilands, J. B. (1982) *Annu. Rev. Microbiol.* **36**, 285–309
3. McIntosh, M. A., and Earhart, C. F. (1977) *J. Bacteriol.* **131**, 331–339
4. Pierce, J. R., Pickett, C. L., and Earhart, C. F. (1983) *J. Bacteriol.* **155**, 330–336
5. Langman, L., Young, I. G., Frost, G. E., Rosenberg, H., and Gibson, F. (1972) *J. Bacteriol.* **112**, 1142–1149
6. Wayne, R., and Neilands, J. B. (1975) *J. Bacteriol.* **121**, 497–503
7. Warner, P. J., Williams, H., Bindereif, A., and Neilands, J. B. (1981) *Infect. Immun.* **33**, 540–545

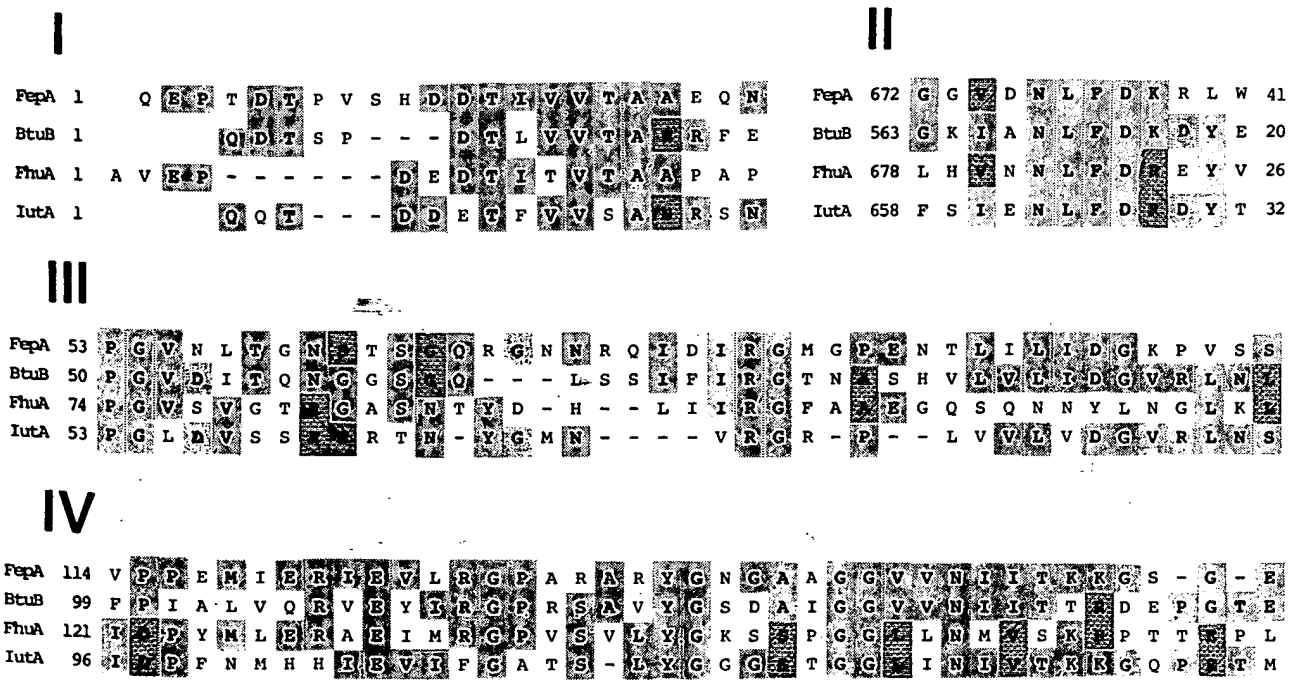

FIG. 3. **Homologous peptide sequences found in FepA, BtuB, FhuA, and IutA.** Spaces were introduced into some sequences to achieve maximum fit. In some cases two shades of gray are used to show that two proteins have 1 amino acid, whereas the other two proteins have a different residue at that position. The *number* on the *left* of the sequences is the number of residues from the mature amino terminus that the homology region begins and for region II the number on the *right* is the distance to the carboxyl terminus. Sequence data for *btuB*, *fhuA*, and *iutA* are from Refs. 12, 34, and 35, respectively.

8. Braun, V. (1981) *FEMS Microbiol. Lett.* **11,** 225–228
9. Fecker, L., and Braun, V. (1983) *J. Bacteriol.* **156,** 1301–1314
10. DeVeaux, L. C., and Kadner, R. J. (1985) *J. Bacteriol.* **162,** 888–896
11. Di Masi, D. R., White, J. C., Schnaitman, C. A., and Bradbeer, C. (1973) *J. Bacteriol.* **115,** 506–513
12. Heller, K., and Kadner, R. J. (1985) *J. Bacteriol.* **161,** 904–908
13. Norrander, J., Kempe, T., and Messing, J. (1983) *Gene (Amst.)* **26,** 101–106
14. Kieny, M. P., Lathe, R., and Lecocq, J. P. (1983) *Gene (Amst.)* **26,** 91–99
15. Coderre, P. E., and Earhart, C. F. (1984) *FEMS Microbiol. Lett.* **25,** 111–116
16. Katz, L., Kingsbury, D. T., and Helinski, D. R. (1973) *J. Bacteriol.* **114,** 577–591
17. Cohen, S. N., Chang, A. C. Y., and Hsu, L. (1972) *Proc. Natl. Acad. Sci. U. S. A.* **69,** 2110–2114
18. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
19. Staden, R. (1980) *Nucleic Acids Res.* **8,** 3673–3694
20. Henikoff, S. (1984) *Gene (Amst.)* **28,** 351–359
21. Fleming, T. P., Nahlik, M. S., and McIntosh, M. A. (1983) *J. Bacteriol.* **156,** 1171–1177
22. Michaelis, S., and Beckwith, J. (1982) *Annu. Rev. Microbiol.* **36,** 435–465
23. Hantke, K. (1981) *Mol. Gen. Genet.* **182,** 288–292
24. de Lorenzo, V., Bindereif, A., Paw, B. H., and Neilands, J. B. (1986) *J. Bacteriol.* **165,** 570–578
25. Rosenberg, M., and Court, D. (1979) *Annu. Rev. Genet.* **13,** 319–353
26. Morgan, W. D., Bear, D. G., Litchman, B. L., and von Hippel, P. H. (1985) *Nucleic Acids Res.* **13,** 3739–3754
27. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132
28. Mizuno, T., Chou, M.-Y., and Inouye, M. (1983) *J. Biol. Chem.* **258,** 6932–6940
29. Gouy, M., and Gautier, C. (1982) *Nucleic Acids Res.* **10,** 7055–7074
30. Grosjean, H., and Fiers, W. (1982) *Gene (Amst.)* **18,** 199–209
31. Dayhoff, M. (1978) *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, National Biomedical Research Foundation, Silver Spring, MD.
32. Lipman, D. J., and Pearson, W. R. (1985) *Science* **227,** 1435–1441
33. Nikaido, H., and Wu, H. C. P. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 1048–1052
34. Coulton, J. W., Mason, P., Cameron, D. R., Carmel, G., Jean, R., and Rode, H. N. (1986) *J. Bacteriol.* **165,** 181–192
35. Krone, W. J. A., Stegehuis, F., Koningstein, G., van Doorn, C., Roosendaal, B., de Graaf, F. K., and Oudega, B. (1985) *FEMS Microbiol. Lett.* **26,** 153–161
36. Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., and Gralla, J. (1973) *Nat. New Biol.* **246,** 40–41

SUPPLEMENTAL MATERIAL TO

Nucleotide Sequence of the Ferri-enterochelin Receptor FepA: Homology Among *Escherichia coli* Outer Membrane Receptors Which Interact with TonB.

by

Michael D. Lundrigan and Robert J. Kadner

TABLE 1. A common sequence in the upstream region of *fepA*, *btuB*, *fhuA*, and *intA*. The number above the last T is the distance from the translation start codon in nucleotides.

	-102	
fepA	TGACTGCGT	
	-161	
btuB	TGATTGCGT	
	-97	
fhuA	AGATTGCGT	
	-144	
intA	TAATCGCGT	

Fig. 1. Alternative stem-and-loop structures in the terminator region of *fepA*. A free energy of -11.4 kcal and -20.4 kcal was calculated for structures I and II respectively by the method of Tinoco et al. (36).

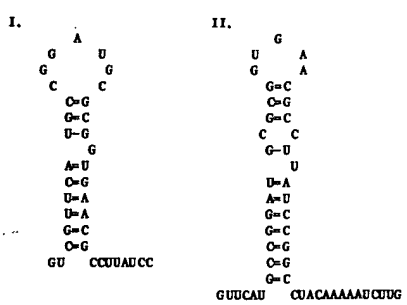

TABLE 2. Amino acid composition of the FepA polypeptide. Parentheses indicate the amino acid compositions of the signal sequence.

Amino Acid	Number of Residues
Ala	50 (3)
Arg	39
Asn	56 (2)
Asp	45
Cys	2
Gln	31 (1)
Glu	39
Gly	74 (2)
His	10 (1)
Ile	35 (2)
Leu	50 (4)
Lys	31 (2)
Met	11 (1)
Phe	17
Pro	29
Ser	48 (1)
Thr	65
Trp	20
Tyr	31 (1)
Val	40 (2)

Number of Residues	723
Molecular Weight	79,908
% Charged Residues	21
Net Charge	-14
Mean Hydropathy	-0.64

TABLE 3. Codon usage for the *fepA* gene. Parentheses indicate the codons used for the signal sequence.

UUU 3	UCU 0	UAU 22 (1)	UGU 0
UUC 14	UCC 7 (1)	UAC 9	UGC 2
UUA 5	UCA 3	UAA 0	UGA 1
UUG 8 (2)	UCG 11	UAG 0	UGG 20
CUU 1	CCU 4	CAU 4 (1)	CGU 23
CUC 4	CCC 0	CAC 6	CGC 11
CUA 0	CCA 6	CAA 7	CGA 2
CUG 32 (2)	CCG 19	CAG 24 (1)	CGG 3
AUU 19 (2)	ACU 7	AAU 12 (1)	AGU 8
AUC 15	ACC 38	AAC 44 (1)	AGC 19
AUA 1	ACA 1	AAA 22	AGA 0
AUG 11 (1)	ACG 19	AAG 9 (2)	AGG 0
GUU 13	GCU 3	GAU 28	GGU 24
GUC 6 (1)	GCC 14 (1)	GAC 17	GGC 37
GUA 5 (1)	GCA 10	GAA 28	GGA 5
GUG 16	GCG 23 (2)	GAG 11	GGG 8 (2)

AUGUST 15, 1986 VOLUME 261 NUMBER 23

ISSN 0021-9258
JBCHA3 261(23) 10467-10952 (1986)

THE Journal of Biological Chemistry

Published by The American Society of Biological Chemists, Inc.

FOUNDED BY CHRISTIAN A HERTER

AND SUSTAINED IN PART BY THE CHRISTIAN A HERTER MEMORIAL FUND